Filtering the Wright-Fisher diffusion.

MIREILLE CHALEYAT-MAUREL[1], VALENTINE GENON-CATALOT[2]

[1]*(Corresponding author) Laboratoire MAP5, Université Paris Descartes, U.F.R. de Mathématiques et Informatique, CNRS UMR 8145 and Laboratoire de Probabilités et Modèles Aléatoires (CNRS-UMR 7599),*
45, rue des Saints-Pères, 75270 Paris Cedex 06, France.
e-mail: mcm@math-info.univ-paris5.fr.
[2]*Laboratoire MAP5, Université Paris Descartes, U.F.R. de Mathématiques et Informatique, CNRS-UMR 8145,*
45, rue des Saints-Pères, 75270 Paris Cedex 06, France.
e-mail: genon@math-info.univ-paris5.fr

Abstract

We consider a Wright-Fisher diffusion $(x(t))$ whose current state cannot be observed directly. Instead, at times $t_1 < t_2 < \ldots$, the observations $y(t_i)$ are such that, given the process $(x(t))$, the random variables $(y(t_i))$ are independent and the conditional distribution of $y(t_i)$ only depends on $x(t_i)$. When this conditional distribution has a specific form, we prove that the model $((x(t_i), y(t_i)), i \geq 1)$ is a computable filter in the sense that all distributions involved in filtering, prediction and smoothing are exactly computable. These distributions are expressed as finite mixtures of parametric distributions. Thus, the number of statistics to compute at each iteration is finite, but this number may vary along iterations.

MSC: Primary 93E11, 60G35; secondary 62C10.

Keywords: Stochastic filtering, partial observations, diffusion processes, discrete time observations, hidden Markov models, prior and posterior distributions.

Running title: Wright-Fisher diffusion.

1 Introduction

Consider a large population composed of two types of individuals A and a. Suppose that the proportion $x(t)$ of A-type at time t evolves continuously according to the following stochastic differential equation

$$dx(t) = [-\delta x(t) + \delta'(1 - x(t))]dt + 2[x(t)(1 - x(t))]^{1/2}dW_t, \quad x(0) = \eta, \tag{1}$$

where (W_t) is a standard one-dimensional Brownian motion and η is a random variable with values in $(0, 1)$ independent of (W_t). This process is known as the Wright-Fisher gene frequency diffusion model with mutation effects. It has values in the interval $(0, 1)$. It appears as the diffusion approximation of the discrete time and space Wright-Fisher Markov chain and is used to model the frequency of an allele A in a population of genes composed of two distinct alleles A and a (see *e.g.* Karlin and Taylor (1981, p. 176-179 and 221-222) or Wai-Yuan (2002, Chap. 6)). Suppose now that the current state $x(t)$ cannot be directly observed. Instead, at times $t_1, t_2, \ldots, t_n$ with $0 \leq t_1 < t_2 \ldots < t_n$, we have observations $y(t_i)$ such that, given the whole process $(x(t))$, the random variables $y(t_i)$ are independent and the conditional distribution of $y(t_i)$ only depends on the corresponding state variable $x(t_i)$. More precisely, we consider the following discrete conditional distributions. Either, a binomial distribution, *i.e.*, for $N \geq 1$ an integer,

$$P(y(t_i) = y|x(t_i) = x) = \binom{N}{y} x^y (1 - x)^{N-y}, \quad y = 0, 1, \ldots, N, \tag{2}$$

or, a negative binomial distribution, *i.e.*, for $m \geq 1$ an integer,

$$P(y(t_i) = y|x(t_i) = x) = \binom{m + y - 1}{y} x^m (1 - x)^y, \quad y = 0, 1, 2, \ldots. \tag{3}$$

Under these assumptions, the joint process $(x(t_n), y(t_n))$ is a hidden Markov model (see *e.g.* Cappé et *al.*, 2005).

In this context, a central problem that has been the subject of a huge number of contributions is the problem of filtering, prediction or smoothing, *i.e.* the study of the conditional distributions of $x(t_l)$ given $y(t_n), \ldots, y(t_1)$, with $l = n$ (filtering), $l = n + 1, n + 2, \ldots$ (prediction), $l < n$ (smoothing). These distributions are generally called filters (respectively exact, prediction or marginal smoothing filters). Although they may be calculated recursively by explicit algorithms, iterations become rapidly intractable and exact formulae are difficult to obtain. To overcome this difficulty, authors generally try to find a parametric family $\mathcal{F}$ of distributions on the state space of $(x(t_n))$ (*i.e.* a family of distributions specified by a finite fixed number of real parameters) such that if $\mathcal{L}(x_0) \in \mathcal{F}$, then, for all n, l, $\mathcal{L}(x(t_l)|y(t_n), \ldots, y(t_1))$ belongs to $\mathcal{F}$. For such models, the term finite-dimensional filters is usually employed. This situation is illustrated by the linear Gaussian Kalman filter (see *e.g.* Cappé *et al.* 2005). There are few models satisfying the same properties as the Kalman filter: It is rather restrictive to impose a parametric family with a fixed number of parameters (see Sawitzki (1981); see also Runggaldier and Spizzichino (2001)).

Recently, new models where explicit computations are possible and which are not finite-dimensional filters have been proposed (see Genon-Catalot (2003), and Genon-Catalot and

Kessler (2004)). Moreover, in a previous paper (Chaleyat-Maurel and Genon-Catalot, 2006), we have introduced the notion of computable filters for the problem of filtering and prediction. Instead of considering a parametric class $\mathcal{F}$, we consider an enlarged class built using mixtures of parametric distributions. The conditional distributions are specified by a finite number of parameters, but this number may vary according to n, l. Still, filters are computable explicitly. We give sufficient conditions on the transition operator of $(x(t))$ and on the conditional distribution of $y(t_i)$ given $x(t_i)$ to obtain such kind of filters. In the present paper, we show that these conditions are satisfied by the model above. Therefore, the conditional distributions of filtering and prediction are computable and we give the exact algorithm leading to these distributions. Moreover, we obtain the marginal smoothing distributions which are also given by an explicit and exact algorithm.

The paper is organized as follows. In Section 2, we briefly recall some properties of the Wright-Fisher diffusion. In Section 3, we recall the filtering-prediction algorithm and the sufficient conditions of Chaleyat-Maurel and Genon-Catalot (2006) to obtain computable filters. Section 4 contains our main results. We introduce the class $\bar{\mathcal{F}}_f$ composed of finite mixtures of parametric distributions fitted to the model (see (16)). We prove that the sufficient conditions hold for this class and give the explicit formulae for the up-dating and the prediction operator (Proposition 4.1, Theorem 4.2, Proposition 4.5). The result concerning the prediction operator is the most difficult part and requires several steps. Then, we turn back in more details to the filtering-prediction algorithm (Proposition 4.6). Moreover, we give the exact distribution of $(y(t_i), i = 1, \dots, n)$ which is also explicit. Hence, if δ, δ' are unknown and are to be estimated from the data set $(y(t_i), i = 1, \dots, n)$, the exact maximum likelihood estimators of these parameters can be computed. At last, in Subsection 4.4, we study marginal smoothing. We recall some classical formulae for computing the marginal smoothing distributions. These formulae involve the filtering distributions, that we have obtained in the previous section, and complementary terms that can be computed thanks to Theorem 4.2. In the Appendix, some technical proofs and auxiliary results are gathered.

2 Properties of the Wright-Fisher diffusion model.

In order to exhibit the adequate class of distributions within which the filters evolve, we need to recall some elementary properties of model (1). The scale density is given by:

$$s(x) = \exp\left(-(1/2)\int^x \frac{-\delta u + \delta'(1-u)}{u(1-u)} du\right) = x^{-\delta'/2}(1-x)^{-\delta/2}, \quad x \in (0,1).$$

It satifies $\int_0 s(x)dx = \infty = \int^1 s(x)dx$ if and only if $\delta \geq 2$ and $\delta' \geq 2$, conditions that we assume from now on. The speed density is equal to $m(x) = x^{\delta'/2-1}(1-x)^{\delta/2-1}$, $x \in (0,1)$. Therefore, the unique stationary distribution of (1) is the Beta distribution with parameters $\delta'/2, \delta/2$ which has density

$$\pi(x) = \frac{x^{\frac{\delta'}{2}-1}(1-x)^{\frac{\delta}{2}-1}}{B(\frac{\delta'}{2}, \frac{\delta}{2})} 1_{(0,1)}(x). \tag{4}$$

For simplicity, we assume that the instants of observations are equally spaced with sampling interval Δ, *i.e.* $t_n = n\Delta$, $n \geq 1$. Hence, the process $(X_n := x(t_n))$ is a time-homogeneous Markov chain. We denote by $p_\Delta(x, x')$ its transition density and by P_Δ its transition operator. The transition density is not explicitly known. However, it has a precise spectral expansion (see *e.g.* Karlin and Taylor, 1981, p.335-336: Note that $2x(t) - 1$ is a Jacobi diffusion process). The results we obtain below are linked with this spectral expansion although we do not use it directly (see the Appendix).

3 Sufficient conditions for computable filters.

First, we focus on filtering and prediction and we consider case (2) with $N = 1$ for the conditional distributions of $Y_i := y(t_i)$ given $X_i = x(t_i)$, *i.e.*, we consider a Bernoulli conditional distribution. The other cases can be easily deduced afterwards (see the Appendix). Let us set

$$P(Y_i = y | X_i = x) = f_x(y) = x^y (1-x)^{1-y}, y = 0, 1, x \in (0, 1). \tag{5}$$

We consider, on the finite set $\{0, 1\}$, the dominating measure $\mu(y) = 1, y = 0, 1$. Thus, $f_x(y)$ is the density of $\mathcal{L}(Y_i | X_i = x)$ with respect to μ.

3.1 Conditional distributions for filtering, prediction and statistical inference.

Denote by

$$\nu_{l|n:1} = \mathcal{L}(X_l \mid Y_n, \ldots, Y_1), \tag{6}$$

the conditional distribution of X_l given $(Y_n, Y_{n-1}, \ldots, Y_1)$. For $l = n$, this distribution is called the optimal or exact filter and is used to estimate the unobserved variable X_n in an *on-line* way. For $l = n + 1$, the distribution is called the prediction filter and is used to predict X_{n+1} from past values of the Y_i's. For $1 \leq l < n$, it is a marginal smoothing distribution and is used to estimate past data or to improve estimates obtained by exact filters.

It is well known that the exact and prediction filters can be obtained recursively (see *e.g.* Cappé *et al.* (2005)). First, starting with $\nu_{1|0:1} = \mathcal{L}(X_1)$, we have

$$\nu_{n|n:1}(dx) \propto \nu_{n|n-1:1}(dx) f_x(Y_n). \tag{7}$$

Hence,

$$\nu_{n|n:1} = \varphi_{Y_n}(\nu_{n|n-1:1}) \tag{8}$$

is obtained by the operator φ_y with $y = Y_n$ where, for ν a probability on $(0, 1)$, $\varphi_y(\nu)$ is defined by:

$$\varphi_y(\nu)(dx) = \frac{f_x(y)\nu(dx)}{p_\nu(y)}, \quad \text{with} \quad p_\nu(y) = \int_{(0,1)} \nu(d\xi) f_\xi(y). \tag{9}$$

This step is the up-dating step which allows to take into account a new observation. Then, we have the prediction step

$$\nu_{n+1|n:1}(dx') = \int_{(0,1)} \nu_{n|n:1}(dx) p_\Delta(x, x') dx' = \nu_{n|n:1} P_\Delta(dx'), \tag{10}$$

which consists in applying the transition operator: $\nu \to \nu P_\Delta$. These properties are obtained using that the joint process (X_n, Y_n) is Markov with transition $p_\Delta(x_n, x_{n+1}) f_{x_{n+1}}(y_{n+1}) dx_{n+1} \mu(dy_{n+1})$, and initial distribution $\nu_{1|0:1}(dx_1) f_{x_1}(y_1) \mu(dy_1)$. Moreover, the conditional distribution of Y_n given $(Y_{n-1}, \ldots, Y_1)$ has a density with respect to μ, given by

$$p_{n|n-1:1}(y_n) = p_{\nu_{n|n-1:1}}(y_n) = \int_{(0,1)} \nu_{n|n-1:1}(dx) f_x(y_n). \tag{11}$$

Note that (Y_n) is not Markov and that the above distribution effectively depends on all previous variables. For statistical inference based on $(Y_1, \ldots, Y_n)$, the exact likelihood is given by

$$p_n(Y_1, \ldots, Y_n) = \prod_{i=1}^{n} p_{\nu_{i|i-1:1}}(Y_i). \tag{12}$$

3.2 Sufficient conditions for computable filters.

Now, we recall the sufficient conditions of Chaleyat-Maurel and Genon-Catalot (2006). First, consider a parametric class $\mathcal{F} = \{\nu_\theta, \theta \in \Theta\}$ of distributions on $(0, 1)$, where Θ is a parameter set included in $\mathbb{R}^p$, such that:

- C1: For $y = 0, 1$, for all $\nu \in \mathcal{F}$, $\varphi_y(\nu)$ belongs to $\mathcal{F}$, *i.e.* $\varphi_y(\nu_\theta) = \nu_{T_y(\theta)}$, for some $T_y(\theta) \in \Theta$.
- C2: For all $\nu \in \mathcal{F}$, $\nu P_\Delta = \sum_{\lambda \in \Lambda} \alpha_\lambda \nu_{\theta_\lambda}$ is a finite mixture of elements of the class $\mathcal{F}$, *i.e.* Λ is a finite set, $\alpha = (\alpha_\lambda, \lambda \in \Lambda)$ is a mixture parameter such that, for all λ, $\alpha_\lambda \geq 0$ and $\sum_{\lambda \in \Lambda} \alpha_\lambda = 1$, and $\theta_\lambda \in \Theta$, for all $\lambda \in \Lambda$.

Proposition 3.1. *Consider now the extended class $\bar{\mathcal{F}}_f$ composed of finite mixtures of distributions of $\mathcal{F}$. Then, under (C1)-(C2), the operators φ_y, $y = 0, 1$ and $\nu \to \nu P_\Delta$ are from $\bar{\mathcal{F}}_f$ into $\bar{\mathcal{F}}_f$. Therefore, once starting with $\nu_{1|0:1} = \mathcal{L}(X_1) \in \bar{\mathcal{F}}_f$, all the distributions $\nu_{n|n:1}$ (exact filters) and $\nu_{n+1|n:1}$ (prediction filters) belong to $\bar{\mathcal{F}}_f$.*

For all n, these distributions are completely specified by their mixture parameter and the finite set of distributions involved in the mixture. Of course, the number of components may vary along the iterations, but still remains finite. Thus, these distributions are explicit and we say that filters are computable.

The proof of Proposition 3.1 is elementary (see Theorem 2.1, p.1451, Chaleyat-Maurel and Genon-Catalot, 2006, see also Proposition 4.6 below). Now, it has a true impact because the extended class is considerably larger than the initial parametric class. Evidently, the difficulty is to find models satisfying these conditions. Examples are given in Chaleyat-Maurel and Genon-Catalot (2006): The hidden Markov process $(x(t))$ is a radial Ornstein-Uhlenbeck process and two cases of conditional distributions of $y(t_i)$ given $x(t_i)$ are proposed. Other examples are given in Genon-Catalot (2003) and Genon-Catalot and Kessler (2004). Here, we study a new and completely different model. It has the noteworthy feature that, to compute filters, the explicit formula for the transition density of $(x(t))$ is not required, contrary to the examples of the previous papers. (Details on the transition density are given in the Appendix).

4 Main results.

Our main results consists in exhibiting the proper parametric class of distributions on $(0, 1)$ and in checking (C1)-(C2) for this class and the model specified by (1) and (5). The interest of these conditions is that they can be checked separately. Condition (C1) only concerns the conditional distributions of Y_i given X_i and the class $\mathcal{F}$. In the Appendix, we prove condition (C1) for the model specified by (1) and (2) or (3). Condition (C2) concerns the transition operator of $(x(t))$. It is the most difficult part. Let us define the following class of distributions indexed by $\Theta = \mathbb{N} \times \mathbb{N}$:

$$\mathcal{F} = \{\nu_{i,j}(dx) \propto h_{i,j}(x)\pi(x)dx, (i, j) \in \mathbb{N} \times \mathbb{N}\}, \tag{13}$$

where

$$h_{i,j}(x) = x^i(1-x)^j. \tag{14}$$

Hence, each distribution in $\mathcal{F}$ is a Beta distribution with parameters $(i + \frac{\delta'}{2}, j + \frac{\delta}{2})$ and (see (4))

$$\nu_{i,j}(dx) = \frac{x^{i+\frac{\delta'}{2}-1}(1-x)^{j+\frac{\delta}{2}-1}}{B(i+\frac{\delta'}{2}, j+\frac{\delta}{2})} 1_{(0,1)}(x)dx. \tag{15}$$

Let us define the extended class:

$$\bar{\mathcal{F}}_f = \{\sum_{(i,j)\in\Lambda} \alpha_{i,j}\nu_{i,j}, \Lambda \subset \mathbb{N}\times\mathbb{N}, |\Lambda| < +\infty, \alpha = (\alpha_{i,j}, (i, j) \in \Lambda) \in S_f\}, \tag{16}$$

where

$$S_f = \{\alpha = (\alpha_{i,j}, (i, j) \in \Lambda), \Lambda \subset \mathbb{N}\times\mathbb{N}, |\Lambda| < +\infty, \forall (i, j), \alpha_{i,j} \geq 0, \sum_{(i,j)\in\Lambda} \alpha_{i,j} = 1\} \tag{17}$$

is the set of finite mixture parameters. It is worth noting that the stationary distribution $\pi(x)dx = \nu_{0,0}(dx)$ belongs to $\mathcal{F}$. Thus, in the important case where the initial distribution, *i.e.* the distribution of η (see (1)), is the stationary distribution, the exact and optimal filters have an explicit formula.

4.1 Condition (C1): Conjugacy.

Proposition 4.1. *Let $\nu_{i,j} \in \mathcal{F}$ (see (13)).*

1. *For $y = 0, 1$, $\varphi_y(\nu_{i,j}) = \nu_{i+y,j+1-y}$. Hence, (C1) holds.*
2. *The marginal distribution is given by*

$$p_{\nu_{i,j}}(y) = \left(\frac{i + \frac{\delta'}{2}}{i + j + \frac{\delta'+\delta}{2}}\right)^y \left(\frac{j + \frac{\delta}{2}}{i + j + \frac{\delta'+\delta}{2}}\right)^{1-y}, y = 0, 1. \tag{18}$$

Proof. The first point is obtained using that $f_x(y)\nu_{i,j}(dx) \propto x^{i+y+\frac{\delta'}{2}-1}(1-x)^{j+1-y+\frac{\delta}{2}-1}dx \propto h_{i+y,j+1-y}(x)\pi(x)dx$.

For the marginal distribution, we have

$$p_{\nu_{i,j}}(y) = \frac{B(i + y + \frac{\delta'}{2}, j + 1 - y + \frac{\delta}{2})}{B(i + \frac{\delta'}{2}, j + \frac{\delta}{2})}.$$

To get (18), we use the classical relations $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma(a+1) = a\Gamma(a)$, $a, b > 0$ where $\Gamma(.)$ is the standard Gamma function. $\square$

Proposition 4.1 is proved for the conditional distributions (2) and (3) in the Appendix.

Remark : Proposition 4.1 states that the class $\mathcal{F}$ is a conjugate class for the parametric family of distributions on $\{0, 1\}$: $y \to f_x(y)\mu(dy)$ with respect to the parameter $x \in (0, 1)$. The property that Beta distributions are conjugate with respect to Bernoulli distributions is well known in Bayesian statistics (see *e.g.* West and Harrison (1997)). Analogously, the class $\mathcal{F}$ is also a conjugate class for the parametric family of distributions (2) and (3) with respect to $x \in (0, 1)$. $\Diamond$

4.2 Condition (C2): Introducing mixtures.

The class $\mathcal{F}$ is the natural class of distributions to consider for the conditional distributions (5) because this class contains the stationary distribution $\pi(x)dx = \nu_{0,0}(dx)$ of (1). We do have $\nu_{0,0}P_\Delta = \nu_{0,0}$. However, when $(i,j) \neq (0,0)$, $\nu_{i,j}P_\Delta$ no more belongs to $\mathcal{F}$ but belongs to $\bar{\mathcal{F}}_f$ as we prove below. We need some preliminary properties.

Proposition 4.2. *For all Borel positive function h defined on $(0,1)$, if $\nu(dx) = h(x)\pi(x)dx$, then, $\nu P_t(dx') = P_t h(x')\pi(x')dx'$, where, for $t \geq 0$,*

$$P_t h(x') = \int_0^1 h(x)p_t(x', x)dx.$$

Proof. It is well known that one-dimensional diffusion processes are reversible with respect to their speed density: The transition $p_t(x,x')$ is reversible with respect to $\pi(x)dx$, *i.e.* satisfies for all $(x,x') \in (0,1)\times(0,1)$:

$$\pi(x)p_t(x,x') = \pi(x')p_t(x',x). \tag{19}$$

This gives the result. $\square$

Proposition 4.3. *Suppose that:*

- *(C3) For all $(n,p) \in \mathbb{N}\times\mathbb{N}$, there exists a set $\Lambda_{n,p} \subset \mathbb{N}\times\mathbb{N}$ such that $|\Lambda_{n,p}| < +\infty$ and for all $t \geq 0$,*

  $$P_t h_{n,p}(.) = \sum_{(i,j)\in\Lambda_{n,p}} B_{i,j}(t)h_{i,j}(.), \tag{20}$$

  with, for all (i,j), and all $t \geq 0$, $B_{i,j}(t) \geq 0$.

Then, condition (C2) holds for P_t for all $t \geq 0$. Moreover,

$$\nu_{n,p}P_t(dx) = \sum_{(i,j)\in\Lambda_{n,p}} \alpha_{i,j}(t)\nu_{i,j}(dx),$$

where $\alpha(t) = (\alpha_{i,j}(t),(i,j)\in\Lambda_{n,p})$ belongs to S_f (see (17)) and

$$\alpha_{i,j}(t) = B_{i,j}(t)\frac{B(i+\frac{\delta'}{2}, j+\frac{\delta}{2})}{B(n+\frac{\delta'}{2}, p+\frac{\delta}{2})}. \tag{21}$$

Proof. We have (see (15)), for all(n,p),

$$\nu_{n,p}P_t = \frac{B(\frac{\delta'}{2},\frac{\delta}{2})}{B(n+\frac{\delta'}{2}, p+\frac{\delta}{2})}(h_{n,p}\pi)P_t.$$

Using Proposition 4.2, we get:

$$(h_{n,p}\pi)P_t(dx) = P_t h_{n,p}(x)\pi(x)dx.$$

Now,

$$P_t h_{n,p}(.)\pi(.) = \sum_{(i,j)\in\Lambda_{n,p}} B_{i,j}(t)h_{i,j}(.)\pi(.),$$

and

$$h_{i,j}(x)\pi(x)dx = \nu_{i,j}(dx)\frac{B(i+\delta'/2, j+\delta/2)}{B(\delta'/2,\delta/2)}.$$

Joining all formulae, we get the result. $\square$

Therefore, it remains to prove condition (C3). We start with a classical lemma.

Lemma 4.1. *Let h belong to the set $C_b^2((0,1))$ of bounded and twice continuously differentiable functions on $(0,1)$. Then,*

$$\frac{d}{dt}(P_t h(x)) = P_t L h(x), P_0 h(x) = h(x), \tag{22}$$

where $Lh(x) = 2x(1-x)h''(x) + [-\delta x + \delta'(1-x)]h'(x)$ is the infinitesimal generator of (1).

Proof. Let $x_x(t)$ be the solution of (1) with initial condition $x_x(0) = x$. By the Ito formula,

$$h(x_x(t)) = h(x) + \int_0^t Lh(x_x(s))ds + 2\int_0^t h'(x_x(s))\,(x_x(s)(1-x_x(s)))^{1/2}\,dW_s.$$

By the assumption on h, taking expectations yields:

$$P_t h(x) = h(x) + \int_0^t P_s L h(x) ds,$$

which is the result. $\square$

Now, we start to compute $P_t h_{n,p}(x)$ for all $n, p \in \mathbb{N}$.

Proposition 4.4. *Let $m_{n,p}(t,.) = P_t h_{n,p}(.)$ Then, for all $n, p \in \mathbb{N}$,*

$$\frac{d}{dt} m_{n,p}(t,.) = -a_{n+p} m_{n,p}(t,.) + c_n(\delta') m_{n-1,p}(t,.) + c_p(\delta) m_{n,p-1}(t,.), \quad m_{0,0}(t,.) = 1, \tag{23}$$

where, for all $n \in \mathbb{N}$,

$$a_n = n[2(n-1) + \delta + \delta'], \quad c_n(\delta) = n[2(n-1) + \delta]. \tag{24}$$

(If p or n is equal to 0, then, $c_p(\delta) = 0$ or $c_n(\delta') = 0$, and formula (23) holds). Note that, since the expression of a_n is symetric with respect to δ, δ', we do not mention the dependance on these parameters. Note also that for all n, since both δ and δ' are positive (actually ≥ 2), the coefficients a_n and $c_n(\delta), c_n(\delta')$ are non negative.

Proof. In view of (22), it is enough to prove that

$$Lh_{n,p} = -a_{n+p} h_{n,p} + c_n(\delta') h_{n-1,p} + c_p(\delta) h_{n,p-1}, \tag{25}$$

where L is defined in Lemma 4.1. To make the proof clear, let us start with computing $Lh_{n,0}$. We have immediately:

$$Lh_{n,0}(x) = -a_n h_{n,0}(x) + c_n(\delta') h_{n-1,0}(x).$$

Now, since $y(t) = 1 - x(t)$ satifies

$$dy(t) = [-\delta' y(t) + \delta(1-y(t))]dt + 2(y(t)(1-y(t))^{1/2} dW_t, \tag{26}$$

we obtain $Lh_{0,p}$ by simply interchanging δ and δ' and get

$$Lh_{0,p}(x) = -a_p h_{0,p}(x) + c_p(\delta) h_{0,p-1}(x).$$

Finally, to compute $Lh_{n,p}$, we use the following tricks: Each time x^{n+1} appears, we write $x^{n+1} = -(1-x-1)x^n = -(1-x)x^n + x^n$; each time $(1-x)^{p+1}$ appears, we write $(1-x)^{p+1} = (1-x)^p(1-x) = (1-x)^p - x(1-x)^p$. Grouping terms, we get (25). $\square$

Our aim is now to prove that

$$m_{n,p}(t,.) = \exp(-a_{n+p}t)h_{n,p}(.) + \sum_{0\le k\le n, 0\le l\le p, (k,l)\neq(0,0)} B^{n,p}_{n-k,p-l}(t)h_{n-k,p-l}(.), \tag{27}$$

where, for all $(k,l),(n,p)$, $B^{n,p}_{n-k,p-l}(t) \ge 0$ for all $t \ge 0$. Moreover, we give below the precise formula for these coefficients. Hence, the set $\Lambda_{n,p}$ of (C3) is equal to $\{(k,l), 0 \le k \le n, 0 \le l \le p\}$. The first term can also be denoted by

$$B^{n,p}_{n,p}(t) = \exp(-a_{n+p}t).$$

It has a special role because it is immediately obtained by (23).

4.2.1 Computation of $m_{n,0}(t,.)$ and $m_{0,n}(t,.)$.

Recall notation (14) and that $m_{n,0}(t,.) = P_t h_{n,0}(.)$. We prove now that (27) holds for all $(n,0)$ and all $(0,n)$.

Theorem 4.1. *The following holds:*

$$m_{n,0}(t,.) = \exp(-a_n t)h_{n,0}(.) + \sum_{k=1}^{n} B^{n,0}_{n-k,0}(t)h_{n-k,0}(.), \tag{28}$$

where, for all (k,n), with $1 \le k \le n$, $B^{n,0}_{n-k,0}(t) \ge 0$ for all $t \ge 0$. Moreover, for $k = 1,\dots,n$,

$$B^{n,0}_{n-k,0}(t) = c_n(\delta')c_{n-1}(\delta')\dots c_{n-k+1}(\delta')B_t(a_n, a_{n-1},\dots,a_{n-k}), \tag{29}$$

where

$$B_t(a_n, a_{n-1},\dots,a_{n-k}) = (-1)^k \sum_{j=0}^{k} \exp(-a_{n-j}t)\frac{(-1)^j}{\prod_{0\le l\le k, l\neq j}|a_{n-j}-a_{n-l}|}. \tag{30}$$

We can also set

$$B^{n,0}_{n,0}(t) = \exp(-a_n t).$$

Analogously:

$$m_{0,n}(t,.) = \exp(-a_n t)h_{0,n}(.) + \sum_{k=1}^{n} B^{0,n}_{0,n-k}(t)h_{0,n-k}(.), \tag{31}$$

where, for all (k,n), with $1 \le k \le n$, $B^{0,n}_{0,n-k}(t) \ge 0$ for all $t \ge 0$. Moreover, for $k = 1,\dots,n$,

$$B^{0,n}_{0,n-k}(t) = c_n(\delta)c_{n-1}(\delta)\dots c_{n-k+1}(\delta)B_t(a_n, a_{n-1},\dots,a_{n-k}),$$

We also set

$$B^{0,n}_{0,n}(t) = \exp(-a_n t).$$

Proof. For the proof, let us fix x and set $m_{n,0}(t,x) = m_n(t)$. We also set $B^{n,0}_{n-k,0}(t) = B^n_{n-k}(t)$ during this proof. Solving $m'_n(t) = -a_n m_n(t) + c_n(\delta')m_{n-1}(t), m_n(0) = x^n = h_{n,0}(x)$ yields

$$m_n(t) = \exp(-a_n t)\, x^n + \exp(-a_n t) \int_0^t \exp(a_n s) m_{n-1}(s) ds. \tag{32}$$

Let us first prove by induction that

$$m_n(t) = \sum_{k=0}^{n} B^n_{n-k}(t) x^{n-k}, \tag{33}$$

where $B^n_{n-k}(t) \geq 0$ for all $t \geq 0$ and all $k = 0, \dots, n$ and $B^n_n(t) = \exp(-a_n t)$. For $n = 0$, $m_0(t) = 1$. For $n = 1$, we solve (32) and get

$$m_1(t) = \exp(-a_1 t)x + c_1(\delta')\frac{(1 - \exp(-a_1 t))}{a_1}.$$

So, (33) holds for $n = 1$ with

$$B^1_1(t) = \exp(-a_1 t), \quad B^1_0(t) = c_1(\delta')\frac{(1 - \exp(-a_1 t))}{a_1} \geq 0. \tag{34}$$

Suppose (33) holds for $n-1$. We now apply (32). Identifying the coefficients of x^{n-k}, $0 \leq k \leq n$, we get:

$$B^n_n(t) = \exp(-a_n t),$$

and for $k = 0, 1, \dots n-1$,

$$B^n_{n-(k+1)}(t) = c_n(\delta') \exp(-a_n t) \int_0^t \exp(a_n s) B^{n-1}_{n-1-k}(s) ds. \tag{35}$$

Hence, (33) holds for all $n \geq 0$ with all coefficients non negative.

Now, we prove (29)-(30) by induction using (35). For $n = 1$, we look at (34) and see that, since $a_0 = 0$,

$$B^1_0(t) = c_1(\delta')(-1)[\frac{\exp(-a_1 t)}{a_1 - a_0} + \frac{(-1)\exp(-a_0 t)}{|a_0 - a_1|}] = c_1(\delta')B_t(a_1, a_0).$$

Now, suppose we have formulae (29)-(30) for $n-1$ and $k = 0, 1, \dots, n-1$. We know that $B^n_n(t) = \exp(-a_n t)$. Let us compute, for $k = 0, 1, \dots, n-1$, $B^n_{n-(k+1)}(t)$ using (35). We have:

$$B^n_{n-(k+1)}(t) = c_n(\delta')c_{n-1}(\delta') \dots c_{n-k}(\delta')(-1)^k \times B, \tag{36}$$

with

$$B = \sum_{j=0}^{k} \exp(-a_n t) \int_0^t \exp((a_n - a_{n-1-j})s) ds \frac{(-1)^j}{\prod_{0 \leq l \leq k, l \neq j} |a_{n-1-j} - a_{n-1-l}|}. \tag{37}$$

Integrating, we get:

$$B = \sum_{j=0}^{k} \exp(-a_{n-1-j} t) \frac{(-1)^j}{(a_n - a_{n-1-j}) \prod_{0 \leq l \leq k, l \neq j} |a_{n-1-j} - a_{n-1-l}|} + (-\exp(-a_n t))A,$$

with

$$A = \sum_{j=0}^{k} \frac{(-1)^j}{(a_n - a_{n-1-j}) \prod_{0 \le l \le k, l \ne j} |a_{n-1-j} - a_{n-1-l}|}. \tag{38}$$

Hence,

$$B = \sum_{j'=1}^{k+1} \exp(-a_{n-j'}t) \frac{(-1)^{j'-1}}{\prod_{0 \le l' \le k+1, l' \ne j'} |a_{n-j'} - a_{n-l'}|} + (-\exp(-a_n t))A. \tag{39}$$

In view of (29)-(30)-(36)-(38)-(39), to complete the proof of (28), it remains to show the following equality:

Lemma 4.2.

$$\sum_{j=0}^{k} \frac{(-1)^j}{(a_n - a_{n-1-j}) \prod_{0 \le l \le k, l \ne j} |a_{n-1-j} - a_{n-1-l}|} = \frac{1}{(a_n - a_{n-1})(a_n - a_{n-2}) \dots (a_n - a_{n-k-1})}.$$

This lemma requires some algebra and its proof is postponed to the Appendix. At last, to get (31), we just interchange δ' and δ in all formulae because of (26). □

4.2.2 Computation of $m_{n,p}(t,.)$.

Recall that $h_{n,p}(x) = x^n(1-x)^p$ and $m_{n,p}(t,.) = P_t h_{n,p}(.)$. Now, we focus on formula (23). It is easy to see that, since we have computed $m_{n,0}(t,.)$ for all n and $m_{0,p}(t,.)$ for all p, then, we deduce from (23) $m_{n,p}(t,.)$ for all (n,p). This is done as follows. Suppose we have computed all terms $m_{i,j-i}(t,.)$ for $0 \le i \le j \le n$, then, we obtain all terms $m_{i,j-i}(t,.)$ for $0 \le i \le j \le n+1$. Indeed, the extra terms are:

- $m_{n+1,0}(t,.)$ that we know already,
- $m_{i,n+1-i}(t,.)$ for $0 < i < n+1$ that is calculated from

  $$\frac{d}{dt} m_{i,n+1-i}(t,.) = -a_{n+1} m_{i,n+1-i}(t,.) + c_i(\delta') m_{i-1,n+1-i}(t,.) + c_{n+1-i}(\delta) m_{i,n+1-i-1}(t,.),$$

- at last, $m_{0,n+1}(t,.)$ that we know already.

This is exactly filling in a matrix composed of the terms $m_{n,p}(t,.)$. Having the first line $m_{0,n}(t,.)$ and the first column $m_{n,0}(t,.)$, we get each new term $m_{i,j}(t,.)$ from the one above $(m_{i-1,j}(t,.))$ and the one on the left $(m_{i,j-1}(t,.))$.

Now, we proceed to get formula (27).

Theorem 4.2. *For all (i,j) such that $0 \le i \le j \le n$,*

$$m_{i,j-i}(t,.) = \exp(-a_j t) h_{i,j-i}(.) + \sum_{0 \le k \le i, 0 \le l \le j-i, (k,l) \neq (0,0)} B^{i,j-i}_{i-k,j-i-l}(t) h_{i-k,j-i-l}(.), \quad (40)$$

with

$$B^{i,j-i}_{i-k,j-i-l}(t) = \binom{k+l}{k} c_i(\delta') \dots c_{i-k+1}(\delta') c_{j-i}(\delta) \dots c_{j-i-l+1}(\delta) B_t(a_j, a_{j-1}, \dots, a_{j-(k+l)}),$$

with the convention that, for $k = 0$, there is no term in $c_.(\delta')$ and for $l = 0$, there is no term in $c_.(\delta)$).

Proof. By (29)-(30)-(35), we have proved that

$$B_t(a_{n+1}, a_n, \dots, a_{n-k}) = \exp(-a_{n+1} t) \int_0^t \exp(a_{n+1} s) B_s(a_n, a_{n-1}, \dots, a_{n-k}) ds. \quad (41)$$

Suppose (40) holds for $0 \le i \le j \le n$. Let us compute the extra terms $m_{i,n+1-i}(t,.)$ for $0 < i < n+1$ using their differential equations. We have

$$m_{i,n+1-i}(t,.) = \exp(-a_{n+1} t) h_{i,n+1-i}(.) + A_i(\delta') + B_i(\delta), \quad (42)$$

with

$$A_i(\delta') = \exp(-a_{n+1} t) \int_0^t \exp(a_{n+1} s)\ c_i(\delta') m_{i-1,n+1-i}(s,.) ds, \quad (43)$$

$$B_i(\delta) = \exp(-a_{n+1} t) \int_0^t \exp(a_{n+1} s)\ c_{n+1-i}(\delta) m_{i,n+1-i-1}(s,.) ds. \quad (44)$$

We apply the induction formula and replace $m_{i-1,n+1-i}(s,.)$, $m_{i,n+1-i-1}(s,.)$ by their development. This yields:

$$\begin{aligned} m_{i-1,n+1-i}(s,.) &= \exp(-a_n s) h_{i-1,n+1-i}(.) \\ &+ \sum_{0 \le k' \le i-1, 0 \le l' \le n+1-i, (k',l') \neq (0,0)} B^{i-1,n+1-i}_{i-1-k',n+1-i-l'}(s) h_{i-1-k',n+1-i-l'}(.), \end{aligned}$$

$$\begin{aligned} m_{i,n+1-i-1}(s,.) &= \exp(-a_n s) h_{i,n+1-i-1}(.) \\ &+ \sum_{0 \le k'' \le i, 0 \le l'' \le n+1-i-1, (k'',l'') \neq (0,0)} B^{i,n+1-i-1}_{i-k'',n+1-i-1-l''}(s) h_{i-k'',n+1-i-1-l''}(.). \end{aligned}$$

In (42)-(43)-(44), the coefficient of $h_{i-1,n+1-i}(.)$ obtained by the above relations only comes from:

$$c_i(\delta') \exp(-a_{n+1} t) \int_0^t \exp(a_{n+1} s) \exp(-a_n s) ds.$$

By (41), this term is equal to:

$$c_i(\delta') B_t(a_{n+1}, a_n) = B^{i,n+1-i}_{i-1,n+1-i-0} = \binom{1}{1} c_i(\delta') B_t(a_{n+1}, a_n). \quad (45)$$

Analogously, the coefficient of $h_{i,n+1-i-1}(.)$ comes from:

$$c_{n+1-i}(\delta)\exp\left(-a_{n+1}t\right)\int_0^t \exp\left(a_{n+1}s\right)\exp\left(-a_n s\right)ds.$$

This term is equal to:

$$c_{n+1-i}(\delta)B_t(a_{n+1},a_n) = B^{i,n+1-i}_{i,n+1-i-1} = \binom{1}{0}c_{n+1-i}(\delta)B_t(a_{n+1},a_n). \tag{46}$$

Now, the coefficient of the current term $h_{i-k,n+1-i-l}(.)$ comes from the sum of the following two terms:

$$b_1 = c_i(\delta')\exp\left(-a_{n+1}t\right)\int_0^t \exp\left(a_{n+1}s\right)B^{i-1,n+1-i}_{i-1-(k-1),n+1-i-l}(s)ds, \tag{47}$$

$(i-1-k' = i-k, n+1-i-l' = n+1-i-l$, thus $k' = k-1, l' = l)$ and

$$b_2 = c_{n+1-i}(\delta)\exp\left(-a_{n+1}t\right)\int_0^t \exp\left(a_{n+1}s\right)B^{i,n+1-i-1}_{i-k,n+1-i-l}(s)ds, \tag{48}$$

$(i-k'' = i-k, n+1-i-1-l'' = n+1-i-l$, thus $k'' = k, l'' = l-1)$. Thus,

$$b_1 = c_i(\delta')\exp\left(-a_{n+1}t\right)\int_0^t \exp\left(a_{n+1}s\right)\binom{k+l-1}{k-1}c_{i-1}(\delta')\dots c_{i-1-(k-1)+1}(\delta')\times$$

$$c_{n+1-i}(\delta)\dots c_{n+1-i-l+1}(\delta)B_s(a_n,a_{n-1},\dots,a_{n-(k+l-1)})ds$$

$$= \binom{k+l-1}{k-1}c_i(\delta')c_{i-1}(\delta')\dots c_{i-k+1}(\delta')c_{n+1-i}(\delta)\dots c_{n+1-i-l+1}(\delta)$$

$$\times B_t(a_{n+1},a_n,a_{n-1},\dots,a_{n+1-(k+l)}).$$

And

$$b_2 = c_{n+1-i}(\delta)\exp\left(-a_{n+1}t\right)\int_0^t \exp\left(a_{n+1}s\right)\binom{k+l-1}{k}c_i(\delta')\dots c_{i-k+1}(\delta')\times$$

$$c_{n+1-(i+1)}(\delta)\dots c_{n+1-i-l+1}(\delta)B_s(a_n,a_{n-1},\dots,a_{n-(k+l-1)})ds$$

$$= \binom{k+l-1}{k}c_i(\delta')c_{i-1}(\delta')\dots c_{i-k+1}(\delta')c_{n+1-i}(\delta)\dots c_{n+1-i-l+1}(\delta)$$

$$\times B_t(a_{n+1},a_n,a_{n-1},\dots,a_{n+1-(k+l)}).$$

Now using that $\binom{k+l-1}{k-1} + \binom{k+l-1}{k} = \binom{k+l}{k}$, we finally obtain that $b_1 + b_2$ is exactly equal to the expected term $B^{i,n+1-i}_{i-k,n+1-i-l}(t)$. $\square$

4.2.3 Back to the mixture coefficients.

Now, we use Proposition 4.3 and formula (21) to obtain the mixture coefficients for $\nu_{i,j-i}P_t$ We begin with a lemma.

Lemma 4.3. *For $0 \le i \le n$ and $0 \le j \le p$, we have*

$$\frac{B(i+\frac{\delta'}{2}, j+\frac{\delta}{2})}{B(n+\frac{\delta'}{2}, p+\frac{\delta}{2})} = \frac{\binom{i+j}{j} a_{n+p}a_{n+p-1}\dots a_{i+1+p}a_{i+p}a_{i+p-1}\dots a_{i+j+1}}{\binom{n+p}{p} c_n(\delta')c_{n-1}(\delta')\dots c_{i+1}(\delta')c_p(\delta)c_{p-1}(\delta)\dots c_{j+1}(\delta)}. \tag{49}$$

In the trivial case $(i,j) = (n,p)$, the quotient is equal to 1. In (49), for $j = p$, there is no term in $c_.(\delta)$, and for $i = n$, there is no term in $c_.(\delta')$.

Proof. We use the relation

$$B(a+1, b+1) = \frac{ab}{(a+b+1)(a+b)} B(a,b). \tag{50}$$

Hence,

$$B(n+\frac{\delta'}{2}, p+\frac{\delta}{2}) = \frac{(n-1+\frac{\delta'}{2})(p-1+\frac{\delta}{2})}{(n+p-1+\frac{\delta'+\delta}{2})(n+p-2+\frac{\delta'+\delta}{2})} B(n-1+\frac{\delta'}{2}, p-1+\frac{\delta}{2}).$$

By (24), we have:

$$\frac{a_n}{2n} = n-1+\frac{\delta'+\delta}{2}, \quad \frac{c_n(\delta')}{2n} = n-1+\frac{\delta'}{2}, \quad \frac{c_p(\delta)}{2p} = p-1+\frac{\delta}{2}.$$

Hence,

$$B(n+\frac{\delta'}{2}, p+\frac{\delta}{2}) = \frac{c_n(\delta')}{2n}\frac{c_p(\delta)}{2p}\frac{2(n+p)}{a_{n+p}}\frac{2(n+p-1)}{a_{n+p-1}} B(n-1+\frac{\delta'}{2}, p-1+\frac{\delta}{2}).$$

Iterating downwards yields:

$$\frac{B(i+\frac{\delta'}{2}, j+\frac{\delta}{2})}{B(n+\frac{\delta'}{2}, p+\frac{\delta}{2})} = \frac{(i+j)!}{(n+p)!}\frac{n!}{i!}\frac{p!}{j!}\frac{a_{n+j}a_{n+j-1}\dots a_{i+j+1}}{c_n(\delta')c_{n-1}(\delta')\dots c_{i+1}(\delta')c_p(\delta)c_{p-1}(\delta)\dots c_{j+1}(\delta)}.$$

This gives (49). □

Now, we have the complete formula for $\nu_{i,j-i}P_t$.

Proposition 4.5. *For $0 \le i \le j$, we have*

$$\nu_{i,j-i}P_t = \sum_{k=0,\dots,i, l=0,\dots,j-i} \alpha^{i,j-i}_{i-k,j-i-l}(t)\ \nu_{i-k,j-i-l},$$

where, for $(k,l) \neq (0,0)$,

$$\alpha^{i,j-i}_{i-k,j-i-l}(t) = \frac{\binom{i}{k}\binom{j-i}{l}}{\binom{j}{k+l}} a_j a_{j-1}\dots a_{j-k-l+1} B_t(a_j, a_{j-1}, \dots, a_{j-k-l}). \tag{51}$$

For $(k,l) = (0,0)$,

$$\alpha^{i,j-i}_{i,j-i}(t) = \exp(-a_j t). \tag{52}$$

The proof is straightforward using Proposition 4.3, Theorem 4.2 and the lemma.

Let us now make some remarks concerning the above result. First, note that the mixture coefficients are symetric with respect to δ' and δ. The non symetric part appears in the distributions $\nu_{i-k,j-i-l}$. Another point is that, looking at $\nu_{i,j-i}P_t$, we see that very few mixture coefficients will be significantly non nul. Indeed, they are all composed of sums of rapidly decaying exponentials.

To illustrate our result, let us compute more precisely some terms, *e.g.* $\nu_{1,0}, \nu_{2,0}$. For $n=1$, $\alpha_{1,0}^{1,0}(t) = \exp(-a_1 t)$, $\alpha_{0,0}^{1,0}(t) = a_1 B_t(a_1, a_0) = 1 - \exp(-a_1 t)$ and $a_1 = \delta' + \delta, a_0 = 0$. Hence:

$$\nu_{1,0} = \exp(-(\delta'+\delta)t)\nu_{1,0} + (1 - \exp(-(\delta'+\delta)t))\nu_{0,0}.$$

For $n=2$,

$$\begin{aligned}
\alpha_{2,0}^{2,0}(t) &= \exp(-a_2 t), \\
\alpha_{1,0}^{2,0}(t) &= a_2 B_t(a_2, a_1) = \frac{a_2}{a_2 - a_1}(\exp(-a_1 t) - \exp(-a_2 t)), \\
\alpha_{0,0}^{2,0}(t) &= a_2 a_1 B_t(a_2, a_1, a_0) = \frac{a_1}{a_2 - a_1}\exp(-a_2 t) - \frac{a_2}{a_2 - a_1}\exp(-a_1 t) + 1,
\end{aligned}$$

with $a_2 = 2(2+\delta'+\delta)$, $a_1 = \delta'+\delta, a_0 = 0$. And so on ....

Note also that Proposition 4.5 and result (40) can be explained by spectral properties of the transition operator P_t. Indeed, considered as an operator on the space $L^2(\pi(x)dx)$, it has a sequence of eigenvalues and an orthonormal basis of eigenfunctions. The eigenvalues are exactly the $(\exp(-a_n t), n \geq 0)$. The eigenfunction associated with $\exp(-a_n t)$ is a polynomial of degree n, linked with the n-th Jacobi polynomial with indexes $(\frac{\delta'}{2} - 1, \frac{\delta}{2} - 1)$ (see the Appendix). Thus, each polynomial $h_{i,j}$ has a finite expansion on this eigenfunctions basis. Therefore, $P_t h_{i,j}$ has also a finite expansion on the same basis. However, from these spectral properties, it is not evident to guess the expansion obtained in (40) nor is it to guess that the expansion contains only positive terms that lead to mixture coefficients.

Finally, Proposition 4.5 shows that, for all $t \geq 0$, $\sum_{0\leq k\leq i, 0\leq l\leq j-i} \alpha_{i-k,j-i-l}^{i,j-i}(t) = 1$. This can be checked directly by formulae (51)-(52) (see the Appendix).

4.3 Working the filtering-prediction algorithm and estimating unknown parameters.

We must now illustrate how Proposition 3.1 allows to obtain explicitly the successive distributions of filtering $\nu_{n|n:1}$ and of (one-step) prediction $\nu_{n+1|n:1}$ (see (6)). Suppose that the initial distribution is $\mathcal{L}(X_1) = \nu_{0,0}$, *i.e.* the stationary distribution of $(x(t))$. After one observation Y_1, we have the up-dated distribution $\nu_{1|1:1} = \varphi_{Y_1}(\nu_{0,0}) = \nu_{Y_1,1-Y_1}$. Then, we apply the prediction step to get $\nu_{2|1:1} = \nu_{Y_1,1-Y_1}P_\Delta$. This distribution is obtained by Proposition 4.5:

$$\nu_{Y_1,1-Y_1}P_\Delta = \sum_{0\leq k\leq Y_1, 0\leq l\leq 1-Y_1} \alpha_{Y_1-k,1-Y_1-l}^{Y_1,1-Y_1}(\Delta)\nu_{Y_1-k,1-Y_1-l}. \tag{53}$$

Then, there is another up-dating for Y_2, and another prediction, and so on. To be more precise, let us state a proposition that explains the use of Proposition 3.1.

Proposition 4.6. *Suppose* $\nu = \sum_{0 \le k \le i, 0 \le l \le j} \alpha_{i-k,j-l}\, \nu_{i-k,j-l}$ *is a distribution of* $\bar{\mathcal{F}}_f$.

1. *Then, for* $y = 0, 1$,

$$\varphi_y(\nu) \propto \sum_{0 \le k \le i, 0 \le l \le j} \alpha_{i-k,j-l}\, p_{\nu_{i-k,j-l}}(y)\, \nu_{i+y-k,j+1-y-l},$$

where the marginal distribution $p_{\nu_{i-k,j-l}}(y)$ *is given in (18). Thus,*

$$\varphi_y(\nu) = \sum_{0 \le k \le i+y, 0 \le l \le j+1-y} \hat{\alpha}_{i+y-k,j+1-y-l}\, \nu_{i+y-k,j+1-y-l},$$

where $\hat{\alpha}_{i+y-k,j+1-y-l} \propto \alpha_{i-k,j-l}\ p_{\nu_{i-k,j-l}}(y)$ *for* $k = 0, 1, \ldots, i, l = 0, 1, \ldots, j$ *and* $\hat{\alpha}_{i+y-k,j+1-y-l} = 0$ *otherwise.*

2. *We have:*

$$\nu P_\Delta = \sum_{0 \le \kappa \le i, 0 \le \lambda \le j} \left(\sum_{0 \le k \le \kappa, 0 \le l \le \lambda} \alpha_{i-k,j-l}\, \alpha^{i-k,j-l}_{i-\kappa,j-\lambda}(\Delta) \right) \nu_{i-\kappa,j-\lambda},$$

where the $\alpha^{i-k,j-l}_{i-\kappa,j-\lambda}(\Delta)$ *are given in Proposition 4.5.*

3. *The marginal distribution associated with* ν *is*

$$p_\nu(y) = \sum_{0 \le k \le i, 0 \le l \le j} \alpha_{i-k,j-l}\, p_{\nu_{i-k,j-l}}(y). \tag{54}$$

It is therefore a mixture of Bernoulli distribution (see Proposition 4.1 and formula (18)).

The first part is straightforward. The second part is an application of Proposition 4.5 with an interchange of sums. Thus, the number of components in the successive mixture distributions grows. Indeed, let us compute the number of mixture components for the filtering distributions. For $\nu_{1|1:1}$, we find $(1 + Y_1)(1 + 1 - Y_1)$; the prediction step preserves the number of components. For $\nu_{n|n:1}$ and $\nu_{n+1|n:1}$, the number of components is $(1 + \sum_{i=1}^n Y_i)(1 + n - \sum_{i=1}^n Y_i)$. However, as noted above, very few mixture coefficients will be significantly non nul. It was also the case for the model investigated in Genon-Catalot and Kessler (2004).

Let us notice that the h-step ahead predictive distribution, $\nu_{n+h|n:1}$ is obtained from $\nu_{n|n:1}$ by applying the operator P_Δ^h, *i.e.* $\nu_{n+h|n:1} = \nu_{n|n:1} P_\Delta^h$. Therefore, this distribution stays in the class $\bar{\mathcal{F}}_f$ and has the same number of mixture components as $\nu_{n|n:1}$.

Now, suppose that δ', δ are unknown and that we wish to estimate these parameters using the data set $(Y_1, \ldots, Y_n)$. The classical statistical approach is to compute the corresponding maximum likelihood estimators. This requires the computation of the exact joint density of this data set which gives the likelihood function (see (12)). For general hidden Markov models, the exact formula of this density is difficult to handle since the integrals giving the conditional densities of Y_i given $(Y_{i-1}, \ldots, Y_1)$ are not explicitly computable (see formula (11)). On the contrary, in our model, these integrals are computable by formula (18). Suppose that the initial distribution is the stationary distribution of (1), *i.e.* $\nu_{0,0}$. For $i = 1$, the law of Y_1 has density $p_{\nu_{0,0}}(y_1)$ given by (18): It is a Bernoulli distribution with parameter $\frac{\delta'}{\delta'+\delta}$. Then, for $i \geq 2$, the conditional distribution of Y_i given $Y_{i-1}, \ldots, Y_1$ has density $p_{\nu_{i|i-1:1}}(y_i)$. It is now a mixture of Bernoulli distributions. The exact likelihood is therefore a product of mixtures of Bernoulli distributions.

4.4 Marginal smoothing.

In this section, we compute $\nu_{l|n:1}$ for $l < n$. To simplify notations, denote by $p(x_l|y_n, \ldots, y_1)$ the conditional density of X_l given $Y_n = y_n, \ldots, Y_1 = y_1$, *i.e.* the density of $\nu_{l|n:1}$ taken at $Y_n = y_n, \ldots, Y_1 = y_1$. Analogously, denote by $p(y_i|y_{i-1}, \ldots, y_1)$ the conditional density of Y_i given $Y_{i-1} = y_{i-1}, \ldots, Y_1 = y_1$. We introduce the backward function:

$$p_{l,n}(y_{l+1}, \ldots, y_n; x), \tag{55}$$

equal to the conditional density of $(Y_{l+1}, \ldots, Y_n)$ given $X_l = x$. By convention, we set $p_{n,n}(\emptyset; x) = 1$. Then, the following forward-backward decomposition holds.

Proposition 4.7. *For $l \leq n$,*

$$p(x_l|y_n, \ldots, y_1) = \frac{p(x_l|y_l, \ldots, y_1)}{\prod_{i=l+1}^{n} p(y_i|y_{i-1}, \ldots, y_1)} p_{l,n}(y_{l+1}, \ldots, y_n; x_l) \tag{56}$$

This result is classical and may be found *e.g.* in Cappé *et al.* (2005). Therefore, the smoothing density is obtained using the filtering density that we have already computed. The denominator in (56) is also available. It remains to have a more explicit expression for the backward function (55). The following proposition gives a backward recursion from $l = n - 1$ down to $l = 1$ for computing (55).

Proposition 4.8. *First, for all n,*

$$p_{n-1,n}(y_n; x) = P_\Delta[f_.(y_n)](x). \tag{57}$$

Then, for $l + 1 < n$,

$$p_{l,n}(y_{l+1}, \ldots, y_n; x) = P_\Delta[f_.(y_{l+1}) p_{l+1,n}(y_{l+2}, \ldots, y_n; .)](x) \tag{58}$$

Proof. We use the fact that (X_n, Y_n) is Markov with transition $p_\Delta(x_n, x_{n+1}) f_{x_{n+1}}(y_{n+1})$. Given $X_{n-1} = x$, X_n has distribution $p_\Delta(x, x_n)dx_n$. Hence,

$$p_{n-1,n}(y_n; x) = \int_0^1 p_\Delta(x, x_n) f_{x_n}(y_n) dx_n,$$

which gives (57). Then, for $n \geq l+2$,

$$p_{l,n}(y_{l+1}, \dots, y_n; x)$$

$$= \int_0^1 p_\Delta(x, x_{l+1}) f_{x_{l+1}}(y_{l+1}) \times \prod_{i=l+2}^{n} p_\Delta(x_{i-1}, x_i) f_{x_i}(y_i) dx_{l+1} \dots dx_n$$

$$= \int_0^1 p_\Delta(x, x_{l+1}) f_{x_{l+1}}(y_{l+1}) p_{l+1,n}(y_{l+2}, \dots, y_n; x_{l+1}) dx_{l+1},$$

which gives (58). □

Let us now apply these formulae to our model. We will show briefly that backward functions can be computed by simple application of Theorem 4.2. Indeed, since

$$f_x(y_n) = h_{y_n, 1-y_n}(x),$$

$$p_{n-1,n}(y_n; x) = P_\Delta h_{y_n,1-y_n}(x) = m_{y_n,1-y_n}(\Delta, x),$$

is obtained by Theorem 4.2. Next, we compute

$$m_{y_n,1-y_n}(\Delta, .) \times h_{y_{n-1},1-y_{n-1}}(.),$$

which is a linear combination of $h_{y_{n-1}+y_n-k, 2-y_{n-1}-y_n-l}$ with $0 \leq k \leq y_n, 0 \leq l \leq 1-y_n$ and apply the transition operator P_Δ to get $p_{n-2,n}(y_{n-1}, y_n; x)$. This is again given by Theorem 4.2. By elementary induction, we see that backward functions are explicit.

References

[1] Cappé O., Moulines E. and Rydèn T. (2005). *Inference in hidden Markov models*, Springer.

[2] Chaleyat-Maurel M. and Genon-Catalot V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stoch. Proc. and Applic.* **116**, 1447-1467.

[3] Genon-Catalot V. and Kessler M. (2004). Random scale perturbation of an AR(1) process and its properties as a nonlinear explicit filter. *Bernoulli* **(10)** (4), 701-720.

[4] Karlin S. and Taylor H.M. (1981). *A Second Course in Stochastic Processes*. Academic Press.

[5] Lebedev N.N. (1972). *Special functions and their applications.* Dover publications, inc..

[6] Nikiforov A., Ouvarov V. (1983). *Fonctions spéciales de la physique mathématique.* Editions Mir, Moscou.

[7] Runggaldier W. and Spizzichino F. (2001). Sufficient conditions for finite dimensionality of filters in discrete time: A Laplace transform-based approach. *Bernoulli* **7** (2), 211-221.

[8] Sawitzki G. (1981). Finite dimensional filter systems in discrete time. *Stochastics*, vol. **5**, 107-114.

[9] Wai-Yuan T. (2002). *Stochastic models with applications to genetics, cancers, AIDS and other biomedical systems*. Series on concrete and applicable mathematics, Vol. **4**. World Scientific.

[10] West M., Harrison J. (1997). *Bayesian forecasting and dynamic models*. Second ed., in: Springer series in statistics, Springer.

5 Appendix

5.1 Proof of Lemma 4.2.

Let us write in more details expression (38). We have

$$\begin{aligned}
A \;=\; & \frac{1}{(a_n - a_{n-1})} \times \frac{1}{(a_{n-1} - a_{n-2})(a_{n-1} - a_{n-3}) \dots (a_{n-1} - a_{n-1-k})} \\
& + \frac{(-1)}{(a_n - a_{n-2})(a_{n-1} - a_{n-2})} \times \frac{1}{(a_{n-2} - a_{n-3}) \dots\dots (a_{n-2} - a_{n-1-k})} + \dots \\
& + \frac{(-1)^{j-1}}{(a_n - a_{n-j})(a_{n-1} - a_{n-j}) \dots (a_{n-j+1} - a_{n-j})} \times \frac{1}{(a_{n-j} - a_{n-j-1}) \dots (a_{n-j} - a_{n-k-1})} \\
& + \dots + \frac{(-1)^k}{(a_n - a_{n-1-k})(a_{n-1} - a_{n-1-k})(a_{n-2} - a_{n-1-k}) \dots (a_{n-k} - a_{n-1-k})}
\end{aligned}$$

Now, we set

$$\begin{aligned}
L_0 &= (a_n - a_{n-1})(a_n - a_{n-2})(a_n - a_{n-3}) \dots (a_n - a_{n-k-1}), \\
L_1 &= (a_{n-1} - a_{n-2})(a_{n-1} - a_{n-3}) \dots\dots (a_{n-1} - a_{n-k-1}), \\
L_2 &= (a_{n-2} - a_{n-3})(a_{n-2} - a_{n-4}) \dots (a_{n-2} - a_{n-k-1}), \\
\vdots \;\; & \;\; \vdots \\
L_{k-1} &= (a_{n-k+1} - a_{n-k})(a_{n-k+1} - a_{n-k-1}), \\
L_k &= (a_{n-k} - a_{n-k-1}).
\end{aligned}$$

We must prove that

$$A = \frac{1}{L_0}. \tag{59}$$

For this, we introduce the product $T_n = L_0 L_1 \dots L_k$. Now, we need to prove that $AT_n = L_1 L_2 \dots L_k$. We start to compute AT_n:

$$\begin{aligned} AT_n &= \frac{L_0}{a_n - a_{n-1}} L_2 \dots L_k \\ &+ \frac{(-1) L_0 L_1}{(a_n - a_{n-2})((a_{n-1} - a_{n-2})} L_3 \dots L_k + \dots \\ &+ \frac{(-1)^{j-1} L_0 L_1 \dots L_{j-1}}{(a_n - a_{n-j})(a_{n-1} - a_{n-j}) \dots (a_{n-j+1} - a_{n-j})} L_{j+1} \dots L_k + \dots \\ &+ \frac{(-1)^k L_0 L_1 \dots L_{k-1}}{(a_n - a_{n-k-1})(a_{n-1} - a_{n-k-1}) \dots (a_{n-k+1} - a_{n-k-1})}. \end{aligned}$$

Now, we see that

$$AT_n = P(a_n) \tag{60}$$

where $P(.)$ is a polynomial with degree k. Indeed, in AT_n, the terms containing a_n come only from the terms

$$\frac{L_0}{(a_n - a_{n-j})} = P_j(a_n),$$

where

$$P_1(x) = (x - a_{n-2})(x - a_{n-3}) \dots (x - a_{n-k-1}),$$

$$P_j(x) = (x - a_{n-1})(x - a_{n-2}) \dots (x - a_{n-j+1}) \times (x - a_{n-j-1}) \dots (x - a_{n-k-1}),$$

$$P_{k+1}(x) = (x - a_{n-1})(x - a_{n-2}) \dots (x - a_{n-k}),$$

are all products of k factors of degree 1. Notice that $P_j(x)$ is nul for $x = a_{n-1}, a_{n-2}, \dots, a_{n-j+1}, a_{n-j-1}, \dots, a_{n-k-1}$ and that $P_1(a_{n-1}) = L_1$,

$$P_j(a_{n-j}) = (a_{n-j} - a_{n-1})(a_{n-j} - a_{n-2}) \dots (a_{n-j} - a_{n-j+1}) \times L_j,$$

$$P_{k+1}(a_{n-k-1}) = (a_{n-k-1} - a_{n-1})(a_{n-k-1} - a_{n-2}) \dots (a_{n-k-1} - a_{n-k}).$$

Therefore (see (60))

$$\begin{aligned} P(x) &= P_1(x) L_2 L_3 \dots L_k \\ &\quad - P_2(x) \frac{L_1}{a_{n-1} - a_{n-2}} L_3 \dots L_k + \dots \\ &\quad + (-1)^{j-1} P_j(x) \frac{L_1 L_2 \dots L_{j-1}}{(a_{n-1} - a_{n-j})(a_{n-2} - a_{n-j}) \dots (a_{n-j+1} - a_{n-j})} L_{j+1} \dots L_k + \dots \\ &\quad + (-1)^k P_{k+1}(x) \frac{L_1 \dots L_{k-1}}{(a_{n-1} - a_{n-1-k})(a_{n-2} - a_{n-1-k}) \dots (a_{n-k+1} - a_{n-1-k})} \end{aligned}$$

Now,

$$
\begin{aligned}
P(a_{n-1}) &= P_1(a_{n-1})L_2L_3\dots L_k = L_1L_2L_3\dots L_k\\
P(a_{n-2}) &= -P_2(a_{n-2})\frac{L_1}{(a_{n-1}-a_{n-2})}L_3\dots L_k = -\frac{(a_{n-2}-a_{n-1})}{(a_{n-1}-a_{n-2})}L_1L_2\dots L_k\\
&= L_1L_2\dots L_k\\
\vdots\quad & \quad\vdots\\
P(a_{n-j}) &= (-1]^{j-1}P_j(a_{n-j})\frac{L_1L_2\dots L_{j-1}}{(a_{n-1}-a_{n-j})(a_{n-2}-a_{n-j})\dots(a_{n-j+1}-a_{n-j})}L_{j+1}\dots L_k\\
&= (-1)^{2(j-1)}L_1L_2\dots L_{j-1}L_jL_{j+1}\dots L_k\\
\vdots\quad & \quad\vdots\\
P(a_{n-k-1}) &= (-1)^kP_{k+1}(a_{n-k-1}))\frac{L_1\dots L_{k-1}}{(a_{n-1}-a_{n-1-k})(a_{n-2}-a_{n-1-k})\dots(a_{n-k+1}-a_{n-1-k})}\\
&= (-1)^{2k}L_1L_2\dots L_k.
\end{aligned}
$$

Therefore, $P(x) = L_1L_2\dots L_k$ for the $k+1$ distinct values $x = a_{n-1}, a_{n-2}, \dots, a_{n-k-1}$. Since $P(x)$ is a polynomial of degree k, it is constant equal to $L_1L_2\dots L_k$. In particular,

$$P(a_n) = L_1L_2\dots L_k,$$

which is equivalent to $A = 1/L_0$ (see (59)). So the proof of Lemma 4.2 is complete.

5.2 Binomial or negative binomial conditional distributions.

Proposition 4.1 holds for the other cases given in the introduction. Let $\nu_{i,j}$ belong to $\mathcal{F}$.

- If $f_x(y) = \binom{N}{y}x^y(1-x)^{N-y}$, $y = 0,\dots,N$, then $\varphi_y(\nu_{i,j}) = \nu_{i+y,j+N-y}$ and the marginal distribution is equal to:

$$p_{\nu_{i,j}}(y) = \binom{N}{y}\frac{B(i+y+(\delta'/2), j+N-y+(\delta/2))}{B(i+(\delta'/2), j+(\delta/2))}, y = 0,1,\dots,N \tag{61}$$

- If $f_x(y) = \binom{m+y-1}{y}x^m(1-x)^y$, $y = 0,\dots$, then $\varphi_y(\nu_{i,j}) = \nu_{i+m,j+y}$ and for $y = 0,1,\dots$,

$$p_{\nu_{i,j}}(y) = \binom{m+y-1}{y}\frac{B(i+m+(\delta'/2), j+y+(\delta/2))}{B(i+(\delta'/2), j+(\delta/2))} \tag{62}$$

5.3 Mixture coefficients.

We now check using formula (51) that $\sum_{0\le k\le i, 0\le l\le j-i}\alpha^{i,j-i}_{i-k,j-i-l}(t) = 1$. By interchanging sums and setting $k' = k, l' = k+l$, we first get

$$\sum_{0\le k\le i,0\le l\le j-i}\alpha^{i,j-i}_{i-k,j-i-l}(t) = \sum_{l'=0}^{j}p(i,j-i)a_ja_{j-1}\dots a_{j-l'+1}B_t(a_j,\dots,a_{j-l'}),$$

where

$$p(i, j-i) = \sum_{0 \le k' \le i, 0 \le l'-k' \le j-i} \frac{\binom{i}{k'}\binom{j-i}{l'-k'}}{\binom{j}{l'}}.$$

We recognize the sum of hypergeometric probabilities so that $p(i, j-i) = 1$. There remains to prove that, for all $i \ge 0$,

$$\sum_{0 \le k \le i} \alpha^{i,0}_{i-k,0}(t) = \sum_{k=0}^{i} a_i a_{i-1} \dots a_{i-k+1} B_t(a_i, \dots, a_{i-k}) = 1.$$

We fix i. Looking at (30) and interchanging sums, we have to check that

$$\sum_{j=0}^{i} H_{i-j} \exp\left(-a_{i-j} t\right) = 1,$$

where, for $j = 0, 1, \dots, i$,

$$H_{i-j} = \sum_{k=j}^{i} L^j_k \tag{63}$$

and

$$L^j_k = (-1)^{k+j} \frac{a_i a_{i-1} \dots a_{i-k+1}}{\prod_{0 \le l \le k, l \ne j} |a_{i-j} - a_{i-l}|}.$$

Since $a_0 = 0$ and $H_0 = (-1)^{2j} a_i \dots a_1 / a_i \dots a_1 = 1$, we have $H_0 \exp\left(-a_0 t\right) = 1$. So we must prove that, for all $j = 0, 1, \dots, i-1$, $H_{i-j} = 0$. Denote by D^j_k the denominator of L^j_k:

$$D^j_k = (a_i - a_{i-j}) \dots (a_{i-j+1} - a_{i-j})(a_{i-j} - a_{i-j-1}) \dots (a_{i-j} - a_{i-k}).$$

It is easy to prove by induction (on k) that, for $k = i-1, \dots, j+1$,

$$L'^j_k := L^j_i + L^j_{i-1} + \dots + L^j_k = (-1)^{k+j} \frac{a_i \dots \hat{a}_{i-j} \dots a_{i-k+1}}{D^j_{k-1}},$$

where the notation $\hat{}$ means that the term is absent. The formula for $k = j+1$ yields

$$L'^j_{j+1} = (-1)^{2j+1} \frac{a_i \dots a_{i-j+1}}{D^j_j} = -L^j_j.$$

This gives $H_{i-j} = 0$ (see (63)) for all $j = 0, 1, \dots, i-1$..

5.4 Spectral approach.

The transition density $p_t(x, y)$ of (1) can be expressed using the spectral decomposition of the operator P_t. Consider equation (1) and set $z(t) = 2x(t) - 1$. Then,

$$dz(t) = [-\delta(1 + z(t)) + \delta'(1 - z(t)]dt + (1 - z^2(t))^{1/2} dW_t.$$

This is a Jacobi diffusion process. Let us set

$$\alpha = \frac{\delta}{2} - 1, \beta = \frac{\delta'}{2} - 1. \tag{64}$$

Then, for $n \geq 0$, $u(z) = P_n^{\alpha,\beta}(z)$ with

$$P_n^{\alpha,\beta}(z) = \frac{(-1)^n}{2^n n!}(1-z)^{-\alpha}(1+z)^{-\beta}\frac{d^n}{dz^n}[(1-z)^{n+\alpha}(1+z)^{n+\beta}], \tag{65}$$

is solution of

$$(1-z^2)u'' + [\beta - \alpha - (\alpha+\beta+2)z]u' = -n(n+\alpha+\beta+1)u. \tag{66}$$

The function (65) is the Jacobi polynomial of degree n with indexes (α, β). The sequence $(P_n^{\alpha,\beta}(z), n \geq 0)$ is an orthogonal family with respect to the weight function $\rho(z) = (1-z)^\alpha(1+z)^\beta 1_{(-1,+1)}(z)$. After normalization, it constitutes an orthonormal basis of $L^2(\rho(z)dz)$ (see *e.g.* Lebedev (1972, p.96-97) or Nikiforov and Ouvarov (1983, p.37)). Now, we set $h(x) = u(2x-1)$ in (66) and get:

$$2x(1-x)h'' + [\beta - \alpha - (\alpha+\beta+2)(2x-1)]h' = -2n(n+\alpha+\beta+1)h.$$

Using the relations (64), we obtain:

$$2x(1-x)h'' + [-\delta x + \delta'(1-x)]h' = -n(2(n-1)+\delta+\delta')h.$$

Hence, $Lh == -a_n h$ where L is the infinitesimal generator of (1). For $n \geq 0$, the sequence

$$Q_n(x) = P_n^{\frac{\delta}{2}-1,\frac{\delta'}{2}-1}(2x-1)$$

is the sequence of eigenfunctions of L. The eigenvalue associated with Q_n is $-a_n$. The transition operator P_t has the same sequence of eigenfunctions, and the eigenvalues are $(\exp(-a_n t))$. We have:

$$Q_n(x) = \frac{(-1)^n}{n!}x^{-(\frac{\delta'}{2}-1)}(1-x)^{-(\frac{\delta}{2}-1)}\frac{d^n}{dx^n}[(1-x)^{n+\frac{\delta'}{2}-1}x^{n+\frac{\delta}{2}-1}].$$

Each polynomial Q_n is of the form (see (14))

$$Q_n(x) = \sum_{i=0}^{n} c_{i,n-i}^n h_{i,n-i}(x). \tag{67}$$

And each $h_{i,j-i}$ can be developped as

$$h_{i,j-i} = \sum_{k=0}^{j} d_k^{i,j-i} Q_k,$$

with $d_k^{i,j-i} = c_k^{-1/2}\int_0^1 h_{i,j-i}(x)Q_k(x)\pi(x)dx$ and $c_k = \int_0^1 Q_k^2(x)\pi(x)dx$. Since $P_t Q_k = \exp(-a_k t)Q_k$,

$$P_t h_{i,j-i} = \sum_{k=0}^{j} \exp(-a_k t) d_k^{i,j-i} Q_k.$$

This approach requires the computation of the coordinates $d_k^{i,j-i}$ and of the coefficients $c_{i,k-i}^k$ of (67). Our method gives directly the expression of $P_t h_{i,j-i}$.

Let us notice that the transition density of (1) has the following expression:

$$p_t(x,y) = \pi(y) \sum_{n=0}^{+\infty} \exp(-a_n t) Q_n(x) Q_n(y) c_n^{-1}, \tag{68}$$

as explained in Karlin and Taylor (1981). Therefore, by using the expression (67) and some computations, it is possible to prove that this transition satisfies also condition (T1) of Chaleyat-Maurel and Genon-Catalot (2006). More precisely, this transition can be expressed as an infinite mixture of distributions of the class $\mathcal{F}$.

This property has the following consequence. Suppose that the initial variable in (1) is deterministic $x(0) = x_0$. Then, $x(t_1)$ has distribution $p_{t_1}(x_0, x)$. This distribution belongs to the extended class $\bar{\mathcal{F}}$ composed of infinite mixtures of distributions of $\mathcal{F}$. We can apply our results to the extended class: The filtering, prediction or smoothing distributions all belong to $\bar{\mathcal{F}}$.